DATAFEST AFRICA 2024 DATATHON

IMPROVING ACADEMIC OUTCOME FOR SECONDARY SCHOOL EDUCATION

Team Name: **Granados**

Team Members

1. Emmanuel Ebiendele

2. Omotosho Samuel

Objectives

I. Identify possible problems students might be facing while writing these exams bearing in mind that this would involve starting from the scratch as most schools do not have the foundational data infrastructure to solve this problem.

II. Generate your data through any means based on the needs of your solution while ensuring that your data adequately reflects the state of the African education ecosystem

III. Design an enterprise data solution for the school's data collection, pipelining, warehousing, automation and reporting needs.

IV. Create an optimised model that predicts the likelihood of a student passing or failing their upcoming exam *based on their academic history*. Consider all factors that could affect a student's upcoming result aside from their previous exam scores.

V. Make relevant recommendations to the Stakeholders on how they can help improve the performance of the students based on your solution

OUR SOLUTION

a) Simulate a Dataset Based on Education System in Akure

To simulate a realistic dataset based on the education system in Akure, Nigeria, we designed it to reflect key attributes that affect students' performance in JAMB and WASSCE. The dataset include students' academic data, demographics, environmental factors, and other important metrics based on current trends in secondary education in Nigeria.

 Key Data Points to Include:

1. Student Demographics:

 Age (range: 15-19 years for SS3 students)

 Gender

  Socioeconomic Status (Low, Medium, High)

  Parental Education (None, Primary, Secondary, Tertiary)

2. Academic Data:

  Continuous Assessment Scores (1st term, 2nd term, 3rd term) for core subjects: Mathematics, English, Physics, Chemistry, and a choice of a 5th subject (e.g., Biology, Government, Literature)

  Attendance Percentage

  Final Exam Scores in mock exams leading up to JAMB/WASSCE

3. Environmental Factors:

  Access to Learning Resources (1 = Yes, 0 = No)

Teacher Qualification (Unqualified, Qualified, Highly Qualified)

School Infrastructure Quality (Low, Medium, High)

Access to Electricity (Hours per day)

Classroom Size (Number of students per class)

4.Extracurricular Involvement:

Participation in extracurricular activities (0 = No, 1 = Yes)

Type of extracurricular activities (Sports, Clubs, None)

5. Psychological Factors:

Stress Levels (Low, Medium, High)

Test Anxiety (Low, Medium, High)

Motivation Level (Low, Medium, High)

6. External Exam Results:

JAMB Score (out of 400)

WAEC Score (aggregated score based on 5 subjects)

Predictive Output: Likelihood of passing (Pass/Fail based on score thresholds)

b) **Data Analysis**

Objective: Identify key factors affecting student performance.

Analysis Techniques:

Descriptive statistics to understand distribution (average scores, pass rates).

- Correlation analysis to identify relationships between factors (e.g., study time and performance).

- Regression analysis to predict how different variables influence exam scores.

c) Model Development

Built a machine learning model to predict the likelihood of student passing or failing an upcoming examination

PROPSED SOLUTION TO IMPLEMENT IN IMPROVING SECONDARY SCHOOL EDUCATION

1. Data Collection Plan

Objective: A comprehensive data collection strategy is proposed to capture both students' academic journeys and external factors influencing their performance.

Methods:

 Student Academic Data: The collection of data on exam scores, continuous assessment records, attendance, study habits, subjects enrolled, and time allocation on each subject is recommended.

 Demographic Data: Data on age, gender, socioeconomic background, family education levels, and access to learning resources (e.g., textbooks, internet) should be collected.

Extracurricular Activities: Monitoring participation in clubs, sports, or other activities that contribute to cognitive and social skill development is advised.

Psychological Data: Surveys can be used to assess students' stress levels, motivation, and test anxiety.

Environmental Data: Tracking factors such as infrastructure quality, access to electricity, classroom size, and teacher qualifications is essential.

2.Data Generation Methods:

Surveys and questionnaires should be administered to both students and teachers to gather subjective insights.

Historical academic performance data can be obtained from previous exams and assessments.

Regular monitoring of attendance, participation, and classroom interactions is essential for ongoing insights.

Recommended Tools:

Google Forms or SurveyMonkey for survey administration.

School Management Systems (SMS) for tracking attendance and academic performance.

Custom mobile or web applications could be developed to track students' daily activities and study habits.

3. Data Structuring & Warehouse Design

Objective: A structured database schema is recommended for comprehensive data management, ensuring scalability and efficiency.

Database Schema:

It is proposed to establish tables for Students, Subjects, Exam Results, Continuous Assessments, Attendance, Demographics, Teacher Data, and Environmental Factors .

Relationships such as Many-to-Many between students and subjects, One-to-Many between students and exam results, and One-to-Many between teachers and subjects will be created.

The use of appropriate data types (e.g., text for names, numeric for scores, and boolean for participation) ensures seamless data processing.

Warehouse Tools:

 PostgreSQL or MySQL are recommended for relational data management.

 For real-time data integration, ETL (Extract, Transform, Load) pipelines such as Apache NiFi or Talend can be utilized.

4. Data Analysis

 Objective: An in-depth analysis will be conducted to uncover key factors affecting student performance, enabling informed decision-making.

Analysis Techniques:

Descriptive statistics will provide insights into distributions like average scores and pass rates.

 Correlation analysis will help identify relationships between variables (e.g., study time and performance).

Regression analysis can be employed to predict how factors such as attendance and parental background impact exam scores.

Clustering will be used to group students based on performance patterns and study habits.

Text mining will help analyze feedback forms to identify recurring challenges faced by students.